Platform Cooperativism Consortium

INSTITUTE FOR THE COOPERATIVE DIGITAL ECONOMY

TRUSTWORTHY AI GOVERNANCE IN COOPERATIVES: A MULTIDIMENSIONAL FRAMEWORK FOR FINANCIAL DECISION-MAKING

By Dorleta Urrutia Oñate

Abstract

Oñate's paper develops a Multidimensional Trust Framework for governing AI systems within cooperative organizations, using the Mondragón Cooperative experience as its main reference point. Dorleta Urrutia Oñate argues that cooperatives have distinctive strengths: democratic ownership, distributed accountability, federated governance, and a commitment to member education. These conditions make cooperatives well suited to develop trustworthy AI without outsourcing accountability to external auditors or compliance units. The framework organizes trust across four interdependent dimensions: 1) technical (validation and explainability), 2) organizational (accountability and oversight), 3) social (participation and legitimacy), and 4) institutional (ethics and regulatory alignment). Three financial decision tasks show how this works in practice: cash-flow deviation analysis, collective risk mapping, and social-impact estimation. The examples are situated within the EU AI Act, CSRD, and GDPR. The paper concludes that AI governance does not necessarily require new institutions. Instead, existing cooperative structures can be thoughtfully adapted to meet the demands of algorithmic decision-making.

Keywords: Trustworthy AI, AI governance, cooperative governance, Mondragón Cooperative Experience, algorithmic accountability, democratic oversight,

1. Introduction

Artificial Intelligence (AI) has become a central actor in organizational decision-making. Predictive models and algorithmic management tools increasingly inform processes that, until recently, relied on human judgment—financial forecasting, risk assessment, or recruitment among them. This evolution is technical and institutional: it redefines accountability, redistributes authority, and reshapes how organizations perceive and manage uncertainty.

Yet governance structures have not advanced at the same pace. Many organizations implement AI systems without an explicit set of procedures for accountability or shared oversight, relying on fragmented documentation or informal review. Empirical evidence confirms this gap. According to the Second Diagnostic Report on AI in the Basque Country (BAIC, 2024), only 12% of organizations are currently aligned with the European regulatory framework on AI ethics and governance, suggesting that compliance and accountability practices remain at an early stage of development even in technologically advanced regions. Policy efforts such as the EU Artificial Intelligence Act (AI Act) (European Commission, 2024) and the OECD framework on Trustworthy AI (OECD, 2021) therefore indicate that the challenge is not only technical optimisation; it concerns how explainability, contestability, and democratic control can be preserved as AI systems become integrated in institutional life.

While cooperative systems differ widely across contexts, those based on democratic ownership and deliberative procedures—such as the Mondragón Cooperative Experience (MCE)—offer an institutional configuration that allows oversight and transparency to be exercised collectively. In such settings, existing governance bodies—assemblies, councils, and federations—already manage decisions that link technical and ethical considerations. Where other organizations often struggle to translate ethical guidelines into operational routines, cooperatives already possess formal mechanisms—assemblies, councils, and federations—through which transparency and oversight can be internalized rather than exter-

nally imposed. Cooperatives offer a normative vision and a practical set-up for trustworthy AI: one where accountability is institutional, not outsourced.

At the same time, the broader field of AI governance remains unsettled. International guidelines—such as the OECD AI Principles or the UNESCO Recommendation on the Ethics of AI—offer normative orientation but limited operational guidance. Compliance frameworks, including the AI Act, provide legal clarity yet still depend on organizations' internal capacity to translate procedural requirements into workable practice. Between abstract ethics and regulatory compliance lies a practical space where governance must be organised from within. Understanding how different institutional models can fill that space is an open research question, particularly for organisations that pursue social or cooperative purposes.

This paper approaches that question by examining how cooperative governance can operationalise trustworthy AI within its existing structures of accountability and participation. Our argument is that cooperatives have the institutional conditions to embed technical validation and democratic control into their ordinary procedures, treating trust not as a value to be declared but as a relationship that can be observed and maintained. The analysis develops a Multidimensional Trust Framework, composed of four interrelated dimensions—technical, organizational, social, and institutional—that together describe how trust is produced, verified, and sustained in cooperative contexts.

The discussion proceeds in five steps. Section 2 reviews the labour and institutional frameworks currently shaping algorithmic governance, including trade-union initiatives, regulatory developments, and solidarity-economy experiments in Spain. Section 3 examines how cooperative governance—exemplified by the MCE—provides a mature institutional foundation capable of integrating trustworthy AI as part of its operational culture. Section 4 introduces a Multidimensional Trust Framework derived from ongoing doctoral research on AI governance in cooperatives, outlining its technical, organizational, social, and institutional components. Complementing the conceptual argument, the paper situates the discussion in the financial governance of cooperatives, using three illustrative analytical tasks—cash-flow monitoring, collective-risk mapping, and social-impact reporting—to demonstrate how explainable, task-linked machine-learning models could be organised to enhance transparency and accountability under the EU

AI Act, CSRD, and GDPR. These cases function as conceptual examples, not empirical pilots, helping to connect governance theory with the concrete challenges of financing, coordination, and accountability in cooperative networks. Section 5 consolidates the implications of this analytical exercise and outlines possible directions for future participatory research. The paper concludes by identifying theoretical, practical, and policy insights that emerge when trustworthy AI is understood not as a compliance task but as a collective capacity embedded in democratic governance. Technical and governance terms are defined in Appendix A (Glossary of Terms).

2. Labor Frameworks on Algorithmic Governance

AI governance increasingly unfolds within the terrain of labour relations. The workplace is where automated decisions meet questions of accountability, fairness, and control. Across Europe, unions, regulators, and civic organisations have begun to test ways of bringing algorithmic systems under democratic oversight. Their efforts show that governance is also a political question about how decision-making power is distributed between workers, management, and technology, and about which institutional venues can legitimately adjudicate disputes arising from automated interventions.

2.1 Union-Based Frameworks

European trade unions have become key actors in negotiating the expansion of algorithmic management (Wray, 2025). The European Trade Union Confederation (ETUC) and UNI Global Union now treat algorithmic decision systems as a new frontier for collective bargaining (‘New Studies Show 42 per Cent of Trade Unions Are Already Negotiating on AI and Algorithmic Management across Service Economy’, n.d.). Position papers emphasise meaningful human oversight, fairness, and disclosure, framing these demands as contemporary expressions of long-standing labour rights and due process in employment relations (UNI Global Union, 2024).

Concrete clauses already exist in several jurisdictions. The Italian food-delivery agreement (2023) obliges platforms to disclose the parameters used in task allocation and performance ranking and to prohibit fully automated deactivations (Wray, 2025). Danish logistics agreements establish joint data committees authorised to examine scheduling algorithms and their distributional effects on shifts and overtime. In Spain, the Rider Law (Royal Decree-Law 9/2021) recognises workers' rights to know the criteria that shape algorithmic decisions affecting employment status and conditions. Building on the forthcoming EU Platform Work Directive (2024), unions increasingly claim procedural rights to explanation, consultation, and prior notice before significant algorithmic changes are rolled out (UNI Global Union, 2024).

Unions also invest in technical literacy (Wray, 2025). Training programmes for data stewards and algorithmic auditors — for example, in initiatives linked to UNI Europa — teach participants how to read model documentation, trace data provenance, and request access to raw data under the GDPR (UNI Global Union, 2024). Partnerships with universities and civic labs extend this expertise, using data access and audits as evidence in bargaining: data and audits become evidence through which workers defend wages, safety, working time, and dignity. Importantly, several unions participate in joint innovation pilots with employers to test scheduling or maintenance systems under negotiated safeguards, indicating a shift from opposition to co-governance of AI under collective control (UNI Global Union, 2024).

2.2 Institutional and Regulatory Frameworks

At the same time, international and European institutions are formalising algorithmic governance through regulation. The International Labour Organization (Baiocco et al., 2022) identifies algorithmic management as a defining challenge for the future of work and highlights three safeguards: transparency of logic, accountability of decision-makers, and enforceable rights to human review.

Two European instruments are particularly consequential. First, the EU Platform Work Directive (European Parliament and Council of the European Union, 2024)

introduces rights to algorithmic transparency and human supervision for platform and gig-economy workers, obliging employers to explain automated decision rules, to monitor adverse impacts, and to prevent fully automated dismissals in disciplinary contexts. Second, the EU AI Act (European Commission, 2024) establishes a risk-based framework for high-risk employment applications—recruitment, evaluation, promotion, and task allocation—requiring documented risk management, data-quality protocols, explainability evidence, logging, and traceable oversight. Complementary guidance by the OECD (2021) and UNESCO (2021) promotes fairness, safety, and contestability as measurable dimensions of trustworthy AI and encourages independent assessments where appropriate.

These initiatives establish a legal baseline, yet they act mainly from outside organisations. Workers are often positioned as data subjects with access or rectification rights rather than as participants in designing and monitoring the systems that govern them. Questions also persist about scope (e.g., decision support vs. automated decisions), proportionality of disclosure, and the interface with GDPR safeguards on automated decision-making. This raises a practical question: what internal organisational arrangements can embed algorithmic accountability, aligning compliance obligations with democratic ownership and everyday decision routines?

2.3 Civil Advocacy and Data-Justice Frameworks

Outside formal labour institutions, a dynamic ecosystem of civil and research initiatives has emerged to monitor algorithmic management and defend data justice. Projects such as Fairwork (University of Oxford), Worker Info Exchange (WIE), Éticas Foundation (Barcelona), and the ObservaTAS network (Spain) combine investigative research, participatory audits, and digital-rights advocacy.

Fairwork publishes annual ratings of digital platforms according to principles of fair pay, working conditions, and accountability, illustrating how automated management shapes income volatility and precarity. Worker Info Exchange supports drivers and couriers in exercising data-access rights under the GDPR to reconstruct how dispatch or deactivation decisions occur, sometimes reveal-

ing proxies that correlate with protected characteristics. Éticas Foundation and ObservaTAS develop participatory audit protocols in public-service and media contexts, bringing together affected groups, engineers, and social researchers to assess bias, document institutional impact, and propose corrective measures. Methodologically, such audits blend testimonies, technical inspection, and small-scale data collection, enabling contestation even when corporate disclosure is limited.

Together, these initiatives practice bottom-up accountability. They regard data as a collective resource and promote joint audits where workers and experts co-produce diagnostic evidence. Their continuity, however, often depends on voluntary cooperation from companies or on public-sector openness, which limits scale and follow-through after reports are published. Even so, their methods feed into national and EU debates and help set expectations for transparency, due diligence, and redress mechanisms in algorithmic management.

2.4 Hybrid Frameworks in Spain: From Managerial Adoption to Critical Appropriation

In Spain, cooperative and solidarity-economy organisations have begun to connect AI adoption with democratic governance. Two recent initiatives illustrate complementary orientations: one managerial, the other socio-political.

The Andalusian federation FAECTA (2024) launched a five-phase programme—diagnosis, strategic planning, deployment, training, and evaluation—to integrate AI tools into cooperative management. From almost one hundred applicants, ten cooperatives were selected. The project, co-funded by the regional government, emphasises practical improvement: process optimisation, energy-efficiency analytics, and digital safety. Participants receive hands-on training with generative and predictive tools (e.g., ChatGPT, Gemini, Copilot) treated as aids for reporting, documentation, and decision support. FAECTA thus acts as a capacity-building intermediary, translating AI adoption into structured training, measurable outputs (training hours, workflow indicators), and feedback cycles with cooperative managers.

By contrast, the Red de Economía Alternativa y Solidaria (REAS) frames AI as a socio-ecological and feminist issue. Its publications warn against data extractivism and cultural homogenisation and advocate critical appropriation: smaller, transparent models trained on ethically sourced datasets and supported by cooperative or public infrastructures. Cooperatives such as Tangente, Colectic, and SomConnexió run experimental labs applying AI to communication accessibility, care-work coordination, and energy monitoring, while maintaining collective decision-making over data use and model updates. This approach prioritises explainability and traceability as conditions for legitimacy, not as add-ons.

Overall, these experiences show how federations and networks are moving beyond fear or passive compliance toward participatory experimentation. FAECTA demonstrates how digital innovation can strengthen managerial capacity through structured training and evaluation, whereas REAS delineates ethical and environmental boundaries that responsible AI should respect. Both contribute to an emerging meso-level ecosystem where algorithmic accountability, cooperative values, and technical literacy evolve in tandem and where federations can curate shared artefacts (model cards, data sheets, incident logs) for member organisations.

These initiatives are still pilot-level and depend on project funding and key personnel. The broader challenge is consolidation: how to build participatory oversight into ordinary governance, budget cycles, and audit routines so that accountability persists beyond the life of grants. The MCE offers a particularly relevant context to explore this transition. With more than six decades of democratic ownership, multi-level coordination, and shared accountability, Mondragon possesses a mature institutional architecture capable of integrating trustworthy AI as part of its cooperative identity rather than as a peripheral compliance task. The next section turns to those attributes and examines how they can host a durable model of algorithmic governance.

3. Cooperative Attributes for Trustworthy AI Governance

3.1 Historical and Institutional Overview

The MCE emerged in the Basque Country during the 1950s, a period of industrial restructuring and deep social inequality. Its origins are closely linked to José María Arizmendiarrieta, a priest and educator who promoted technical training and collective ownership as tools for regional development. Arizmendiarrieta treated cooperation as a practical system of industrial democracy—one capable of joining productivity, education, and shared responsibility in a coherent framework (Whyte, 1995).

The first cooperative, ULGOR (later Fagor), was established in 1956 by a group of former students from the technical school founded by Arizmendiarrieta a decade earlier. Its success encouraged the creation of Caja Laboral Popular (today Laboral Kutxa) in 1959, a financial cooperative intended to provide credit and technical support to new ventures (Whyte, 1995). These initial steps defined a model in which financial solidarity and entrepreneurial initiative advanced together, transforming Mondragon into a coordinated ecosystem rather than a collection of isolated enterprises (Imaz, 2023).

By the 1980s, Mondragon had expanded into a diversified network integrating industrial cooperatives, a cooperative bank, educational and research institutions, and welfare mechanisms such as Lagun Aro. This evolution shows an ability to adapt cooperative principles to increasingly complex organizational realities while preserving participatory decision-making (Cheney & Hernandez, 2012). The International Co-operative Alliance (ICA) principles—democratic control, education, inter-cooperation, and concern for community—became progressively integrated in its statutes and management routines.

Over time, these principles turned into an institutional culture rather than a mere legal framework. As Imaz (2023) observes, Mondragon developed a multi-level coordination system where accountability operates through federated governance instead of hierarchical command (Bakaikoa et al., 2004). This structure ena-

bles worker-members to exercise genuine control while allowing the organization to compete globally. Today, the MCE continues to evolve amid technological and economic transitions, reaffirming its founding conviction that cooperation, when organized through well-documented governance, can sustain both competitiveness and social purpose (Mondragón Corporación, 2024) .

3.2 Governance Architecture and Cooperative Principles

Mondragon's governance model represents one of the most structured applications of ICA principles within a large-scale enterprise network. The seven principles—voluntary and open membership, democratic control, member economic participation, autonomy and independence, education, cooperation among cooperatives, and concern for community—function as operational norms, not symbolic declarations (Whyte, 1995).

At the level of each cooperative, governance follows a triadic structure balancing strategic direction, execution, and representation. The Governing Council sets long-term objectives, approves management plans, and oversees implementation. The Management Team conducts day-to-day operations and reports to the Council, remaining accountable to the membership base. The Social Council, distinctive to Mondragon, acts as an internal communication channel between members and management, facilitating dialogue and collective problem-solving through formal consultation procedures (Arando et al., 2011).

All of these bodies operate under the authority of the General Assembly, where every member participates directly under the "one person, one vote" rule. The Assembly approves strategic decisions, elects council members, and validates annual accounts, ensuring that management remains anchored in democratic legitimacy. This architecture distributes power horizontally rather than vertically and builds in mechanisms for transparency and reversibility—core conditions for long-term accountability (Imaz, 2023)

Education plays a pivotal role in sustaining this system. Members receive training in cooperative governance, financial management, and participatory deci-

sion-making, turning the ICA principle of education into a functional requirement for self-management. Autonomy and independence are equally safeguarded: each cooperative governs itself, yet shared statutes and federative agreements delineate the obligations and limits of inter-cooperation (Sunsundegi & Idiakez, n.d.).

Mondragon's system translates cooperative ideals into an organisational constitution—a network of institutions, routines, and checks that make democratic participation compatible with professional management. Originally devised for industrial and financial coordination (Imaz, 2023), this logic now provides a relevant precedent for questions of digital accountability and data governance that similarly demand balances between autonomy, transparency, and collective oversight (Mondragón Corporación, 2022).

3.3 Digital Transition and AI Governance

Over the last decade, digital transformation has become central within the MCE, automation, and the strategic role of data. Rather than separating technological innovation from cooperative governance, Mondragon has worked to integrate digital tools within its institutional and ethical framework. This approach reflects a continuity running throughout its history: adapting new forms of production and management to cooperative principles of participation, transparency, and shared benefit.

The Mondragon White Paper on Artificial Intelligence (2022) articulates this transition explicitly. It presents AI as an opportunity to enhance competitiveness while reaffirming cooperative values the cooperative values that have guided decision-making for decades. The document emphasises that technological development must remain human-centred, transparent, and trustworthy—fully aligned with European AI policy and Mondragon's internal governance culture. It also calls for the creation of shared "infraestructuras cooperativas" for data and AI, designed to allow knowledge and technology to circulate across cooperatives under collective control.

The Socio-Business Policy 2025–2028 consolidates this direction, which designates digital transformation and innovation as cross-cutting priorities. The policy integrates governance objectives that combine "robust processes, clear roles, and relationships of trust" with agility and inter-cooperation, explicitly linking digital innovation to the broader social mission of the cooperative group (Mondragón Corporación, 2024). It identifies ethics, sustainability, and education as the three enabling conditions for responsible digitalization, extending the cooperative emphasis on human development into the digital sphere.

In parallel, the AI & Robotics Community of Mondragon—a network that brings together cooperatives such as IKERLAN, LKS Next, Mondragon Assembly, and Eroski—is developing the technical and organizational capabilities needed for this transition. The community fosters collaboration in industrial automation, predictive maintenance, and energy efficiency, while promoting reflection on explainability, cybersecurity, and algorithmic ethics (Tulankide Magazine, 2024). By coordinating research and technological deployment across firms, it operates as a federated learning environment consistent with Mondragon's distributed-governance ethos.

Together, these initiatives indicate a shift from adoption to governance.. They illustrate how Mondragon's federated structure, educational ecosystem, and ethical orientation are being reinterpreted to address challenges such as data ownership, algorithmic transparency, and responsible innovation. The group's ongoing strategy suggests that the cooperative governance model—rooted in participation and accountability—can provide a practical foundation for trustworthy AI, where democratic oversight and technological sophistication function as mutually reinforcing dimensions of organizational resilience.

3.4 Compliance at Mondragon: Current Practice (AI Act, GDPR, CSRD)

Mondragon cooperatives are already adapting their internal governance systems to meet the requirements of the EU AI Act (2024), the General Data Protection Regulation (GDPR), and the Corporate Sustainability Reporting Directive (CSRD). The group integrates these responsibilities into existing bodies, maintaining the participatory logic that defines its institutional architecture.

At the operational level, model developers and data teams document algorithmic systems using Model Cards, Datasheets for Datasets, and Decision Logs. These artefacts describe each system's purpose, data sources, validation metrics, and usage constraints, translating cooperative norms of transparency and accountability into technical documentation . Oversight is distributed across several layers. The AI and Compliance Lead coordinates documentation and risk registers; the Governing Council authorises high-risk use cases and monitors legal conformity; and the Social Council acts as a mediation channel for member concerns, ensuring that human oversight remains collective rather than individual (Mondragón Corporación, 2022).

Compliance with the AI Act follows a workflow that mirrors pre-existing cooperative review procedures. High-risk applications—those involving employment, finance, or member evaluation—must include: (1) a documented risk-management plan; (2) clear data-governance protocols; (3) evidence of technical robustness and explainability; and (4) human-oversight mechanisms for appeal or reversal (European Commission, 2024). Escalation occurs through a defined pathway: incident reporting at the operational level, triage by the AI Oversight Committee, validation by the Governing Council, and, when appropriate, deliberation in the General Assembly (Mondragón Corporación, 2022).

This distributed workflow builds compliance into cooperative governance and avoids outsourcing it to external auditors. Each cooperative retains autonomy to adapt its procedures, while the federated structure—through the AI & Robotics Community and the Corporate Compliance Office—facilitates shared learning and harmonisation across the group (Mondragón Corporación, 2024). The out-

come is a system of internalised accountability in which regulatory adherence, technical assurance, and democratic participation reinforce one another within a unified governance ecosystem.

3.5 Cooperative Attributes and Implications for AI Governance

The evolution of the MCE illustrates how governance mechanisms originally designed for industrial coordination can also underpin the responsible governance of data-driven and algorithmic systems. Throughout its history, Mondragon has shown that economic organisation can be structured around collective decision-making, shared accountability, and continuous learning—principles that closely align with emerging frameworks for trustworthy and responsible AI.

Several attributes of Mondragon's governance are particularly relevant to the design of AI oversight mechanisms:

First, democratic control guarantees human oversight and contestability. Decision processes are transparent and reversible, with deliberation distributed across multiple bodies rather than concentrated in technical or managerial elites (Arando et al., 2011; Whyte, 1995). This design provides a practical precedent for embedding collective judgment within algorithmic decision-making and for redefining oversight as a shared, rather than individual, responsibility.

Second, distributed accountability parallels algorithmic transparency. The triadic structure of councils and assemblies functions as an internal audit system, ensuring traceability of decisions and collective responsibility for outcomes (Imaz, 2023). These features mirror the procedural safeguards essential to AI governance—comprehensive documentation, systematic monitoring, and the ability to review and correct automated processes.

Third, inter-cooperation offers a model for rethinking data governance. Historical mechanisms of solidarity and resource pooling—such as the Central Inter-Cooperative Fund and Lagun Aro—find contemporary equivalents in shared data infrastructures and the AI & Robotics Community (Tulankide Magazine, 2024). The White Paper on AI (2022) and the Socio-Business Policy 2025–2028 both highlight

cooperation in digital innovation and ethical technology adoption as strategic levers for building collective intelligence and data sovereignty.

Fourth, education and ethical formation sustain organisational adaptability. Mondragon's long-standing commitment to cooperative education now translates into algorithmic literacy—the capacity of members to understand, question, and co-govern digital systems (Sunsundegi & Idiakez, n.d.). The combination of educational programmes, research centres, and applied-innovation units creates a living infrastructure for participatory ethics in technology development.

Finally, social purpose and sustainability remain the normative anchors of innovation. By linking technological progress with environmental responsibility and community benefit, Mondragon articulates a model of purpose-driven digitalisation consistent with the European vision of human-centred AI and ESG-oriented governance (Mondragón Corporación, 2022).

These attributes indicate that the cooperative model puts AI governance into practice through established institutions. Democratic control sustains human oversight. Traceability and explainability follow from distributed accountability. Inter-cooperation enables federated stewardship. Education underpins algorithmic literacy. Social purpose anchors ethical alignment. These continuities suggest that the cooperative governance framework can serve as both a conceptual foundation and a practical testbed for trustworthy AI (Mondragón Corporación, 2022).

This conclusion leads directly to the next section, where these principles are formalised within a multidimensional framework of trust connecting technical validation, organisational design, social participation, and institutional alignment.

4. The Multidimensional Trust Framework

4.1 Purpose and Scope

The Multidimensional Trust Framework proposed in this study is designed to operationalize trustworthy AI within cooperative governance systems such as the MCE. It builds directly on the cooperative attributes identified in the previous section—democratic control, distributed accountability, inter-cooperation, education, and social purpose—and translates them into measurable and governable properties of AI systems. The framework treats trust as a verifiable relationship between models, organisations, and members, expressed through technical validation, procedural accountability, and institutional alignment.

The framework has two aims: (i) it provides a conceptual basis for understanding how trust can be decomposed into interdependent dimensions—technical, organizational, social, and institutional—that together define the reliability and legitimacy of AI systems. (ii) it offers a methodological structure for assessing these dimensions empirically through metrics, documentation, and participatory evaluation. In doing so, it connects abstract principles of trustworthy AI (as articulated in the EU AI Act, 2024 and OECD, 2024 guidelines) with concrete mechanisms of cooperative governance.

The scope is intentionally broad yet applicable to federated organizational contexts. It can guide cooperatives, associations, and social enterprises beginning to integrate AI in management, finance, logistics, or impact assessment, helping them identify where governance gaps persist and how technical practices can reinforce social accountability. By combining regulatory compliance with participatory oversight, the framework aims to support the design, validation, and monitoring of AI systems in environments where transparency and democracy are institutional requirements rather than optional features.

Ultimately, the framework serves as a bridge between trust as a normative ideal and trust as an operational outcome. It shows how cooperative organizations can translate governance culture into technical and procedural safeguards, turning

values such as participation, solidarity, and transparency into measurable components of AI reliability and fairness.

To illustrate the framework's applicability, the discussion refers to three representative analytical tasks that concentrate many of the governance tensions cooperatives encounter around financing and accountability. These are: (i) cash-flow deviation analysis, where explainable forecasting models could support shared financial planning; (ii) collective risk mapping, where interpretable classification tools might help federations detect correlated exposures; and (iii) social-impact estimation, where transparent regression models could link operational data to impact-reporting indicators under the CSRD. Each case functions as a conceptual probe—showing how explainability, traceability, and human oversight can be organised around decision processes that have tangible economic and ethical implications. They are used here as examples to ground the multidimensional framework, not as empirical pilots.

4.2 Structure of the Framework

The Multidimensional Trust Framework is organized around four interdependent dimensions—technical, organizational, social, and institutional—that together define the conditions under which AI systems can be considered trustworthy. Each dimension represents a specific domain of governance and measurement, yet they operate as a continuous system of feedback and mutual reinforcement. The framework is deliberately analytical and diagnostic: it helps identify where trust is generated, where it is at risk, and how governance interventions might restore or strengthen it.

Its arrangement mirrors Mondragon's own governance logic. Just as cooperative governance distributes authority across councils, federations, and assemblies, trust in AI emerges from multiple and interrelated forms of validation. No single layer—technical accuracy, organizational control, social legitimacy, or regulatory compliance—is sufficient on its own. Trust becomes durable when these dimensions interact coherently.

The technical dimension corresponds to transparency and competence. It cap-

tures how the model behaves—its fidelity, robustness, stability, and explainability—and provides quantitative evidence of reliability.

Organisationally, it's built through shared responsibility and traceable oversight. It translates cooperative governance structures into AI oversight mechanisms: defined roles, documentation flows, and escalation procedures that secure human and institutional control over automated processes.

Socially, it grows from collective participation and comprehension.. It concerns the degree to which members and stakeholders understand, trust, and can contest algorithmic decisions; participation here is both cognitive (comprehension of explanations) and procedural (capacity to influence or reverse outcomes).

Institutionally, trust is anchored in cooperative norms and external regulation. It ensures that AI deployment remains consistent with ICA values—education, inter-cooperation, concern for community—as well as with legal and ethical frameworks such as the EU AI Act (2024) and OECD (2024) recommendations.

Each dimension is both independent and relational: technical evidence supports organizational accountability; social legitimacy reinforces institutional compliance; and institutional norms constrain technical design choices. In cooperative contexts, these relationships can be explicitly governed, documented, and evaluated, turning the abstract notion of "trustworthy AI" into a structured governance process.

This approach renders trust operational. It links the cooperative logic of shared responsibility with the technical logic of model validation, allowing organizations like Mondragon to integrate algorithmic systems without compromising their democratic and ethical foundations.

4.3 Technical Dimension: Validation and Explainability

The technical dimension defines trust in terms of how reliably and transparently an AI system performs its intended task. It focuses on measurable properties—fidelity, robustness, stability, and explainability—that describe system behaviour under different conditions. In cooperative settings, these properties are not merely technical indicators; they are forms of evidence that enable accountability and collective understanding. They allow members and decision-makers to evaluate whether outputs align with cooperative values of fairness, transparency, and responsibility.

Technical trust begins with validation, understood as systematic verification that predictions correspond to empirical reality within acceptable margins of uncertainty. Validation includes both classical metrics (accuracy, precision, recall, error rates) and trust-oriented metrics such as calibration, uncertainty quantification, and sensitivity analysis. These metrics can be interpreted as procedural analogues of cooperative auditing—mechanisms for verifying that internal processes meet collectively defined standards of reliability (Lundberg & Lee 2017; Ribeiro et al. 2016).

A second component is explainability, which ensures that the rationale behind decisions can be understood and interrogated. Techniques such as SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), and counterfactual analysis (e.g., DiCE) allow both engineers and non-specialists to observe how input features contribute to outputs. In a cooperative setting, this transparency supports deliberation: members can ask why a model produced a given outcome, and those responsible for governance can adjust rules, thresholds, or datasets accordingly.

The Mondragon White Paper on AI (2022) identifies transparency, interpretability, and reliability as key ethical and operational criteria for AI adoption within the group. It argues that models should be “auditable, explainable, and verifiable,” aligning validation processes with internal principles of quality management and social responsibility (Libro Blanco IA Mondragon 2022). In this sense, the tech-

nical dimension provides the empirical foundation for organizational and social trust: it converts model performance into verifiable evidence that can be examined, discussed, and improved through participatory governance.

4.4 Organizational Dimension: Accountability and Oversight

The organizational dimension concerns how accountability is structured, distributed, and enacted around AI systems. It extends the cooperative principle of shared governance to algorithmic decision-making, ensuring that automated processes remain transparent, reversible, and subject to democratic control. Whereas the technical dimension produces evidence of reliability, the organizational dimension determines who interprets that evidence and how corrective actions are taken when issues arise (Imaz 2023; Arando et al. 2011).

In Mondragon's governance model, accountability is not concentrated in a single authority but distributed across multiple levels—management teams, governing councils, and assemblies. This distribution offers a template for AI oversight: rather than relying on isolated compliance officers or purely technical auditors, responsibility for algorithmic systems can be built within existing deliberative structures. For example, a dedicated AI Oversight Committee could operate analogously to the Social Council, reviewing algorithmic decisions, monitoring bias reports, and validating updates prior to deployment. Such a committee would combine technical expertise with member representation, ensuring that governance reflects both organizational objectives and cooperative values (Política Socioempresarial 2025–2028).

Documentation plays a central role. Model Cards, Datasheets for Datasets, and Decision Logs formalize accountability through traceability. They allow significant changes—parameter adjustments, retraining events, policy modifications—to be recorded, contextualized, and reviewed (EU AI Act 2024). These artefacts are both technical and social: they render algorithmic decisions legible to engineers and auditable by governance bodies. In cooperative terms, they function as the minutes of algorithmic deliberation, connecting day-to-day model management with the broader system of organizational trust.

Escalation protocols put reversibility into practice. When an automated decision generates disagreement, members or managers should have clear channels to request human review or appeal. These processes replicate existing feedback loops in Mondragon's participatory architecture, where conflicts and proposals are mediated between councils and assemblies (Arando et al. 2011; Imaz 2023). By institutionalizing feedback at the algorithmic level, the organization keeps technological systems open to scrutiny and correction rather than allowing them to act as opaque intermediaries of authority.

The Socio-Business Policy 2025–2028 highlights the need for "clear roles and relationships of trust" in digital transformation, calling for governance processes that combine agility with accountability. It understands AI adoption not as a purely technical upgrade but as an organizational redesign requiring ethical oversight and capacity building (Política Socioempresarial 2025–2028). Under this logic, AI oversight becomes a shared managerial function—linking operational reliability (technical trust) to institutional legitimacy (social and regulatory trust).

This organizational dimension transforms cooperative governance structures into the scaffolding of AI accountability. By formalizing roles, documenting processes, and maintaining participatory review mechanisms, cooperatives can ensure that algorithmic systems stay aligned with both performance goals and collective values. Accountability thus appears not only as compliance, but as a procedural expression of the cooperative ethos in the governance of intelligent systems.

4.5 Social Dimension: Participation and Legitimacy

The social dimension situates trust in the interaction between people and algorithms. It addresses how members, workers, and stakeholders understand, evaluate, and influence decisions produced by AI systems. In cooperative contexts, this dimension draws on a long tradition of participatory deliberation: decisions are legitimate not only because they are procedurally correct, but because those

affected can understand, question, and reshape their outcomes (Whyte & Whyte 1991; Arando et al. 2015).

Participation and comprehension are treated as core governance functions, not as after-the-fact communication tasks. For an AI system to be trustworthy within a cooperative, it must generate explanations that are meaningful to non-technical audiences and actionable within existing decision structures. Explainability, in this sense, becomes a dialogical process: it enables members to engage with the system, test assumptions, and deliberate on implications. Transparency shifts from one-way disclosure to an ongoing conversation about behaviour and purpose.

Educational practices are central. Mondragon's commitment to cooperative education—training members in governance, finance, and organizational principles—extends naturally to algorithmic literacy. Just as members learn to read financial reports or production metrics, they can learn to interpret model outputs, assess uncertainty, and recognize potential bias. Workshops, focus groups, and co-design sessions provide opportunities to build shared understanding, making technical systems part of the social learning infrastructure (Libro Blanco IA Mondragon 2022).

Legitimacy grows when participatory mechanisms are routinized. For example, model review assemblies may be scheduled periodically to discuss how algorithmic systems affect workplace coordination, resource allocation, or performance evaluation. These forums mirror the participatory rituals already embedded in cooperative life—spaces where dialogue leads to decision. The result is reinforcement of both individual agency and collective accountability: members not only receive information about AI decisions but also help shape how the systems evolve over time (Freundlich 2017).

The Socio-Business Policy 2025–2028 frames digital transformation as a shared responsibility requiring "learning, dialogue, and ethical reflection," linking technological competence to civic engagement (Política Socioempresarial 2025–2028). It aligns with the cooperative view that technical systems—however advanced—should remain extensions of democratic processes, not substitutes for them.

The social dimension reframes explainability and transparency as collective practices. Trust does not arise solely from technical performance or legal compliance; it grows from members' continuous capacity to comprehend and contest algorithmic decisions. By embedding education, deliberation, and dialogue into AI governance, cooperatives can transform opacity into a new form of participatory literacy—one that makes understanding a condition of legitimacy.

4.6 Institutional Dimension: Ethics and Regulatory Alignment

The institutional dimension locates trust at the intersection of internal cooperative norms and external legal and ethical frameworks. It addresses how organizations ensure that their use of AI complies with regulatory obligations while remaining faithful to foundational values. In cooperative contexts, this dimension formalizes what has long been implicit in Mondragon's governance model: aligning economic activity with social and ethical responsibility (ICA 2023; Imaz 2023).

Institutional trust rests on two parallel forms of accountability. The first is internal, rooted in cooperative statutes and shared ethical commitments—democratic control, education, inter-cooperation, and concern for community—defining legitimate behaviour within the organization. The second is external, referring to compliance with public regulation and societal expectations—what the EU AI Act (2024) describes as "trustworthy, human-centered, and risk-proportionate" governance (EU AI Act 2024). When both operate coherently, ethical reflection and regulatory compliance reinforce each other rather than collide.

The Socio-Business Policy 2025–2028 integrates this dual logic. It defines ethical governance as a strategic pillar, linking integrity, transparency, and sustainability to innovation and digitalization. The policy calls for models that "combine robust processes with clear responsibilities and relationships of trust," effectively translating cooperative principles into organizational norms for technology development (Política Socioempresarial 2025–2028). In this sense, the policy operates as an internal constitution for digital ethics, ensuring that the standards governing financial and social activities extend to data and AI systems.

Externally, the European and international policy landscape provides complementary guidance. The EU AI Act sets procedural obligations—risk management, documentation, human oversight, and transparency—that align with cooperative governance mechanisms already in place. Similarly, the OECD Framework for the Classification of AI Systems (2024) promotes transparency, fairness, and accountability as measurable dimensions of trustworthy AI (OECD 2024). For cooperatives, compliance is less an imposition than a way of making explicit the ethical commitments placed under their institutional DNA.

This alignment also opens room for innovation. By integrating cooperative ethics with formal procedures—AI charters, ethics committees, and audit protocols—cooperatives can contribute to the wider regulatory ecosystem as living laboratories of responsible AI. Their federated structure allows them to act collectively in creating sectoral standards and shared infrastructures for algorithmic accountability, extending inter-cooperation into the normative and policy domain.

The institutional dimension turns ethics and regulation into an operational layer of trust. It ensures that technical validation, organizational oversight, and social participation are embedded within a consistent ethical-legal framework. For Mondragon, this means AI systems are governed not only by compliance requirements but by cooperative principles that have long defined its legitimacy—a model in which prioritise internal review bodies and shared artefacts over standalone compliance units.

4.7 Interactions and Mutual Reinforcement

Trust in AI does not emerge from a single property—accuracy, oversight, participation, or compliance—but from their interaction. The Multidimensional Trust Framework therefore treats trust as a systemic outcome, generated through continuous reinforcement between technical, organizational, social, and institutional mechanisms. This holistic view reflects Mondragon's cooperative logic, where stability and legitimacy depend on balanced coordination across levels rather than top-down command (Imaz 2023; Arando et al. 2015).

Technical reliability provides the base. When models are validated, explainable,

and robust, they produce evidence that governance bodies can examine and act upon. Yet evidence alone is not sufficient; it requires organizational interpretation. Metrics of fidelity or bias matter when included in procedures—audits, review committees, escalation paths—that specify who evaluates them and how they inform decision-making (EU AI Act 2024). This translation from measurement to accountability keeps validation socially and institutionally grounded.

Organizational accountability, in turn, reinforces social legitimacy. When members know there are clear roles, transparent documentation, and accessible channels for contestation, reliance on AI-assisted processes increases. Participatory monitoring—through workshops, assemblies, and feedback loops—turns trust from a passive attitude into a shared practice. This is consistent with democratic member control, where oversight is distributed and collective rather than delegated to a technical elite (Whyte & Whyte 1991; Freundlich 2017).

Social participation also supports institutional alignment. Deliberative processes generate legitimacy that extends beyond the organization, helping cooperatives demonstrate that their AI practices meet internal expectations and external ethical-legal standards. When transparency and education become part of daily governance, compliance shifts from obligation to shared culture (Política Socioempresarial 2025–2028; ICA 2023).

Finally, institutional norms close the loop by stabilizing the other dimensions. Ethical frameworks and regulatory requirements set boundaries for acceptable design and deployment, constraining technical and organizational choices in ways that sustain collective trust. Regulation functions less as an external constraint and more as a feedback mechanism that preserves coherence over time (OECD 2024).

This cyclical interaction—from validation to accountability, from participation to regulation, and back again—allows trust to persist as technologies evolve. In cooperative contexts, many of these feedback loops are already institutionalized through assemblies, federations, and social councils. The challenge, and the opportunity, is to extend those mechanisms to algorithmic systems so that every technical decision remains visible, reversible, and collectively owned.

In this sense, the multidimensional model does not merely describe how trust

can be achieved; it explains how it is maintained. It conceptualizes trust as a dynamic equilibrium among interdependent processes—each providing checks, evidence, and meaning for the others. When these interactions are stable, the system exhibits what might be called organizational resilience through explainability: a condition in which the reliability of technology, the accountability of governance, the legitimacy of participation, and the integrity of ethics sustain one another in continuous balance.

4.8 Illustrative Applications: Financial Decision Tasks in Cooperatives

The multidimensional framework can be illustrated through three representative decision processes commonly found in cooperative finance. These examples are not empirical pilots but analytical mappings that show how participation, validation, institutional compliance, and internal accountability intersect in practice. They serve to translate the framework's abstract logic into the types of governance situations that cooperatives already face in managing financial stability and collective resources.

(1) Cash-flow deviation analysis.

Predictive models are increasingly used to anticipate liquidity gaps and plan working capital. Within a cooperative setting, an explainable regression model could estimate short-term cash-flow deviations, linking each prediction to transparent input factors—revenues, costs, payment cycles. Organisationally, the key issue is who validates and interprets the model's outputs. Technically, reliability depends on calibration and stability. Social trust rests on members' ability to follow the reasoning behind adjustments, while institutionally the process is anchored in internal audit and AI Act documentation. This example shows how algorithmic forecasting could remain auditable and collectively accountable.

(2) Collective risk mapping.

Federated cooperative networks often face correlated exposures across sectors, clients, or supply chains. A transparent classification model could identify shared

risk patterns, informing collective decision-making in risk committees or mutual-aid mechanisms. Here, social trust first takes shape in how federated actors deliberate over mitigation measures. Technical reliability then depends on the robustness of correlation metrics. At the institutional level, the exercise is tied to governance instruments under the CSRD and cooperative statutes, and organisationally it relies on how committees communicate and act upon results. The case illustrates how explainability and inter-cooperation can reinforce accountability beyond individual enterprises.

(3) Social-impact estimation.

Under the CSRD, cooperatives increasingly need to quantify social and environmental outcomes. An explainable model using operational and employment data could estimate impact indicators—such as local procurement rates or job-quality metrics—and express their drivers through SHAP or counterfactual analysis. Institutional alignment guarantees coherence with CSRD and ICA principles. Socially, members deliberate on which indicators are most meaningful. Technical validation secures accuracy and transparency, and organisational routines determine how data are gathered and verified. This example highlights how explainable AI can be used not to optimize performance narrowly, but to support deliberation about social purpose and responsibility.

These examples demonstrate how the four trust dimensions can be analyzed in real decision contexts without requiring full empirical deployment. They turn cooperative finance into a conceptual testing ground for trustworthy AI governance, showing how the same logic that organizes participatory oversight in cooperatives can guide the responsible design and evaluation of algorithmic systems.

5. Conclusion and Future Directions

This paper has developed a conceptual framework that links cooperative governance and AI governance through the lens of trust. It shows how the structures that organize decision-making in cooperatives can inform the supervision and interpretation of AI systems. Rather than proposing a novel institutional blueprint, the analysis has highlighted existing mechanisms—councils, assemblies, federations—that already perform many of the functions currently grouped under the label of "trustworthy AI."

The framework clarifies how layers of trust can be assessed; it does not resolve governance challenges outright. Using the MCE as a reference point, the paper shows that democratic participation and distributed accountability are not abstract commitments but working practices that can sustain responsible technology use. In this sense, the contribution is primarily descriptive: it translates between two domains that seldom interact and provides a shared vocabulary for future work at their intersection.

5.1 Cooperative Governance as Institutional Innovation

Mondragon's system of councils and federations illustrates that governance can be organized around transparency and participation without sacrificing efficiency. This observation matters for AI governance because it suggests that democratic oversight can be maintained even in data-rich, complex environments. The cooperative model demonstrates how decision authority may be distributed, how oversight can be routine rather than exceptional, and how accountability can be shared instead of individualized.

None of these ideas is new—Mondragon has applied them for decades—yet they acquire renewed relevance in the context of algorithmic management and digital decision-making. The argument here is not that cooperatives offer a universal alternative; rather, specific practices—shared auditing, periodic review, member education—can inform the organizational design around AI. The value of

the cooperative example lies in its continuity: a long-standing case where social responsibility, technological adaptation, and internal governance have evolved together.

Viewed this way, Mondragon is less a template to be replicated than a source of practical lessons. It shows how organizations can link control and trust, and how ethics can be upheld through procedure as much as through intent. This approach aligns with ongoing European discussions on data governance and algorithmic oversight, indicating that small, well-defined governance mechanisms can make technological systems more predictable and socially accountable.

5.2 Trust as a Cross-Scale Principle

The analysis suggests that trust can be observed and organized across multiple levels of governance. At the technical level, it concerns model performance and the extent to which behaviour can be validated or explained. At the organizational level, it depends on how responsibilities are distributed and how decisions are documented. Socially, it reflects whether members understand and accept the use of technology. Institutionally, it connects with regulation and shared ethical norms.

These levels are often treated in isolation in debates on AI governance. The framework advanced here treats them as interdependent: each supports and constrains the others. Technical validation has limited value if organizational processes cannot act on its findings, and institutional principles remain aspirational if they are not included in everyday routines. The cooperative model offers a practical illustration of how these layers can be linked through regular procedures—reporting, assembly review, federated coordination—that preserve coherence between local practice and shared standards.

Understanding trust in this way avoids reducing it to either technical performance or moral aspiration. Trust becomes a property of the whole system—distributed, procedural, and open to verification. The perspective does not prescribe a single path for building trust, but it helps identify where it may falter and what kinds of coordination are required to sustain it over time.

5.3 Democratic Implications and Future Directions

The Multidimensional Trust Framework developed in this paper offers a conceptual and analytical base, but its ultimate purpose is empirical: to guide how trust in AI can be observed, measured, and strengthened within cooperative organizations. Future work will operationalize the framework through a Participatory Action Research (PAR) approach, integrating qualitative and quantitative methods that reflect both the technical and social dimensions of trust (Reason & Bradbury 2008).

The research will follow the framework's multidimensional logic. It will begin with diagnostic interviews to map how cooperatives currently deploy—or plan to deploy—AI systems: which data are used, which decisions are automated, and what governance procedures are in place. These interviews will surface existing practices of validation, oversight, and participation, providing a minimum standard for assessing trust.

Subsequent focus groups and co-design workshops will bring together technical staff, managers, and cooperative members to interpret findings and identify governance improvements. These sessions will examine how the four dimensions—technical, organizational, social, and institutional—manifest in daily operations and where gaps or tensions appear. Their participatory design mirrors cooperative deliberation: rather than studying organizations from the outside, the process invites members to reason collectively about how AI affects autonomy, fairness, and efficiency (Arando et al., 2011; Imaz, 2023).

Complementing these qualitative insights, the study will develop a trust metrics dashboard that translates the framework into operational indicators. The tool will document model-level measures (e.g., fidelity, stability, robustness), process indicators (e.g., review frequency, escalation paths), and social indicators (e.g., comprehension, contestability, legitimacy). By integrating these metrics, cooperatives can track the evolution of trust over time, detect early signals of governance risk, and institutionalize corrective actions through their federated structures.

A key feature of the approach will be the feedback mechanism: results are not only collected and analyzed but also returned to participating cooperatives through shared reflection sessions. This ensures that knowledge production remains reciprocal and that the research process contributes directly to organizational learning and governance improvement—a principle consistent with participatory research and with the cooperative ethos of mutual accountability (Imaz, 2023).

The expected outcomes are twofold. First, to test and refine the Multidimensional Trust Framework in real cooperative environments, providing evidence of how trust can be sustained across socio-technical systems. Second, to generate practical tools—templates, metrics, governance procedures—that cooperatives can adapt when adopting AI technologies. Through this process, the framework moves from conceptual model to theory-in-use: a living methodology for governing trustworthy AI in organizations where democracy, solidarity, and transparency are both ethical imperatives and operational realities.

Cooperatives largely rely on AI supplied or hosted by large vendors and rarely control foundation models or cloud infrastructure. Accordingly, the paper focuses on auditable, task-linked machine learning and on governance–compliance mechanisms that organizations can realistically own and improve.

This future direction carries a broader implication. It suggests that AI governance does not depend on creating entirely new institutions, but on adapting and reinforcing existing ones so they can meet the demands of digital transformation. By grounding the study of trust in cooperative practice, the research aims to contribute not only to AI governance as a field but also to a practical understanding of how democratic values can shape technological systems in everyday organizational life.

The paper has outlined a way to connect cooperative governance and AI governance through a common concern: how to sustain trust as decision-making becomes increasingly technical. The Multidimensional Trust Framework and the Mondragon case indicate that many conditions for trustworthy AI—transparency, accountability, participation—already exist in familiar institutional forms. The

task ahead is not to invent mechanisms from scratch but to understand how existing ones can adapt, coordinate, and learn.

Seen from this perspective, cooperatives offer a practical reminder rather than a new model: governance, at its core, is a collective effort to make decisions visible, revisable, and fair. If trust in AI rests on those same capacities, then studying how they are sustained in real organizations becomes a meaningful way forward—pilot quarterly model-review cycles across two cooperatives.

References

Arando, S., Freundlich, F., Gago, M., & Jones, D. C. (2011). Assessing Mondragon: Stability and managed change in the face of globalization. *Organization*, 18(5), 619–639. https://doi.org/10.1177/1350508411414293

BAIC, B. A. I. C. (2024). *Second Diagnostic Report on Artificial Intelligence in the Basque Country*. Basque Artificial Intelligence Center.

Baiocco, S., Fernandez-Macías, E., Rani, U., & Pesole, A. (2022, June 21). *The Algorithmic Management of work and its implications in different contexts | International Labour Organization.*

Bakaikoa, B., Errasti, A., & Begiristain, A. (2004). Governance of the Mondragon Corporación Cooperativa. *Annals of Public & Cooperative Economics*, 75(1), 61–87.

Cheney, G., & Hernandez, A. (2012). *WORKER COOPERATIVE VALUES AND OWNERSHIP CULTURE.*

European Commission. (2024). *Artificial Intelligence Act (EU Regulation 2024/1689)*. Official Journal of the European Union.

European Parliament and Council of the European Union. (2024). D*IRECTIVE (EU) 2024/2831 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL* on improving working conditions in platform work.

Imaz, I. (2023). Governing by participation: Institutional learning in the Mondragon cooperative system. *Journal of Co-Operative Studies*, 56(2), 24–37.

Mondragón Corporación. (2022). *Libro Blanco sobre Inteligencia Artificial en el Grupo Mondragón*. Mondragón Corporación.

Mondragón Corporación. (2024). *Política Socioempresarial del Grupo Mondragón 2025–2028.* Mondragón Corporación.

New studies show 42 per cent of trade unions are already negotiating on AI and algorithmic management across service economy. (n.d.). *UNI Europa*. Retrieved 27 October 2025, from https://www.uni-europa.org/news/new-studies-show-42-

per-cent-of-trade-unions-are-already-negotiating-on-ai-and-algorithmic-management-across-service-economy/

OECD. (2021, June 27). *Tools for trustworthy AI*. OECD. https://www.oecd.org/en/publications/tools-for-trustworthy-ai_008232ec-en.html

Sunsundegi, I. O., & Idiakez, A. L. (n.d.). *PERTSONEN ERANTZUKIZUNA ENPRESA KOOPERATIBOAN*.

Tulankide Magazine. (2024). Comunidad de IA y Robótica de Mondragón: Colaboración y desarrollo ético en inteligencia artificial. *Tulankide*, 758, 34–37.

UNI Global Union. (2024). *ALGORITHMIC MANAGEMENT: OPPORTUNITIES FOR COLLECTIVE ACTION: A GUIDE FOR WORKERS AND TRADE UNIONS*.

Whyte, W. F. (1995). Learning from the Mondragón cooperative experience. *Studies in Comparative International Development*, 30(2), 58–67. https://doi.org/10.1007/BF02802954

Wray, B. (2025, September 23). Negotiating the Algorithm: Trade Union Manual. *Brave New Europe.* https://braveneweurope.com/ben-wray-negotiating-the-algorithm-trade-union-manual

Appendix A – Glossary of Terms

This glossary provides concise definitions of the technical and governance terms used throughout the paper. Each entry indicates the trust dimensions to which it primarily contributes: T = Technical, O = Organizational, S = Social, I = Institutional.

General and Conceptual

Task-linked model — (T/O): A supervised machine-learning model designed for a specific organizational task, with defined inputs, outputs, and metrics aligned to a decision process; prioritizes traceability and auditability over general-purpose flexibility.

Explainability (XAI) — (T/S/O): Set of methods and practices that make a model's input–output logic understandable. Supports deliberation, review, and corrective action.

Transparency vs. Interpretability — (T/S): Transparency refers to information availability (documentation, logs); interpretability refers to the cognitive ability to understand how outputs are produced.

Technical Methods and Metrics

SHAP (Shapley Additive Explanations) — (T/S): Method that attributes an additive contribution of each feature to a prediction, ensuring local and global consistency; useful for interpretability and audit.

LIME (Local Interpretable Model-Agnostic Explanations) — (T/S): Produces local surrogate explanations by approximating the model around a single instance; less stable than SHAP but intuitive for exploration.

Counterfactual explanation — (T/S/O): Describes the minimal input changes needed to obtain a different outcome (e.g., "if receivables dropped by 5%, the risk flag would deactivate"). Supports "what-if" reasoning and managerial action.

Fidelity (of explanations) — (T): The degree to which an explanation accurately reflects the underlying model's behavior. High fidelity = low distortion between explanation and model.

Stability (of explanations) — (T/S): The consistency of explanations under small data or parameter perturbations; crucial for trust and organizational learning.

Robustness (of models) — (T): Resistance of model performance to reasonable variations in data or context (noise, distribution shift); assessed via out-of-sample and stress tests.

Calibration — (T/S): The match between predicted probabilities and observed frequencies (reliability). Important for financial and risk-related decisions.

Uncertainty quantification — (T/O): Measurement of predictive uncertainty (e.g., confidence intervals, variance) used to inform thresholds, rejection options, and escalation.

Sensitivity analysis — (T): Examines how model outputs change when inputs or parameters vary; identifies dominant variables and potential vulnerabilities.

Monotonicity constraint — (T): A modelling restriction ensuring expected relationships (e.g., higher risk → higher risk score), enhancing interpretability and auditability.

Documentation and Oversight Tools

Model Card — (O/S/I): Standardized document describing a model's purpose, data sources, evaluation metrics, limitations, and results.

Datasheet for Datasets — (O/I): Record describing dataset origin, license, representativeness, quality, and known risks.

Decision Log — (O/I): Chronological record of configuration changes, retrainings, thresholds, and responsible parties; supports traceability and review.

Audit trail / Logging — (O/I): Technical and procedural logs allowing reconstruction of a system's lifecycle for auditing and evidence.

Governance and Accountability

Human oversight — (O/I/S): Formal mechanisms for human review, validation, or reversal of AI-assisted decisions.

Contestability — (S/O/I): The right and practical ability to question or appeal algorithmic outcomes and obtain meaningful review.

Escalation path — (O/I): Predefined route for escalating incidents or doubts (e.g., operational → AI committee → governingcouncil → general assembly).

Risk-management plan (AI Act) — (O/I): Documented process for identifying, mitigating, and monitoring risks throughout an AI system's lifecycle.

Data-governance protocol — (O/I): Rules covering data collection, quality, access, retention, and security; defines roles and controls.

Traceability — (O/T/I): Ability to link predictions to specific data sources, model versions, and execution conditions.

Impact indicator (CSRD) — (I/S/O): Reportable metric on social or environmental performance (e.g., quality employment, local procurement), with clear definitions and assurance criteria.

Participatory review assembly — (S/O): A formal and periodic review space where technical staff and member representatives jointly evaluate system behaviour.

Trust dashboard — (T/O/S/I): Consolidated panel integrating technical (fidelity, calibration), organizational (review cadence), social (understanding, legitimacy), and institutional (compliance) indicators.

Rejection option / Defer to human — (T/O/S): Policy to abstain from automated decision-making when uncertainty or fairness risk exceeds thresholds, delegating to human review.

Federated risk mapping — (O/S/I): Coordinated interpretation of risk signals across cooperatives or entities under shared protocols and data-access rules.

Published by the Institute for the Cooperative Digital Economy, April 2026
This research report is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License.

Platform
Cooperativism
Consortium

INSTITUTE
FOR THE
COOPERATIVE
DIGITAL
ECONOMY